Technology Infusion Working Group

Earth Science Data Systems Working Group Meeting
November 2011
Newport News, VA

Working Group Co-Chairs:
Karen Moe, NASA/ESTO
Brian Wilson, NASA/JPL

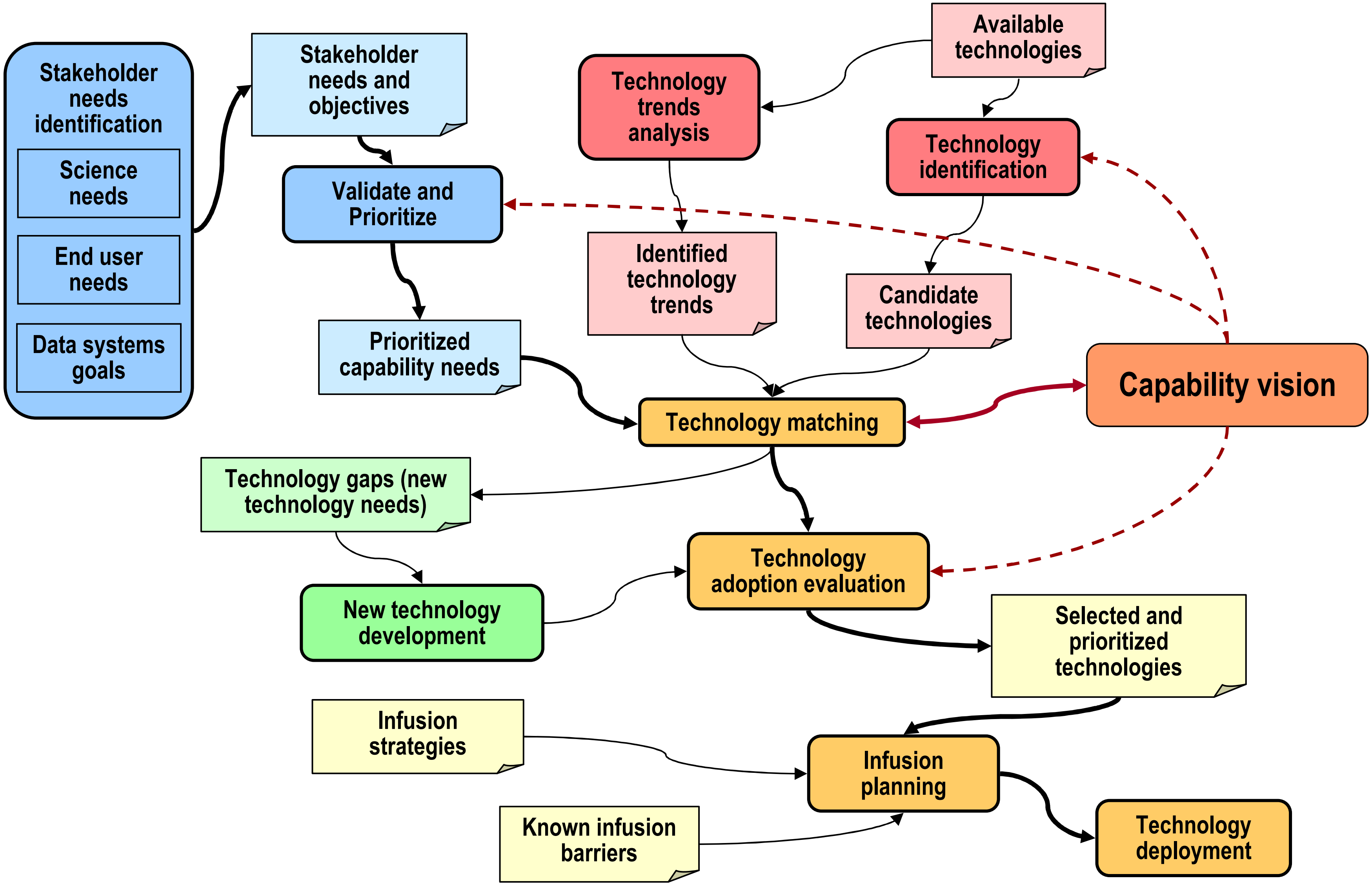
Agenda

- Mission & Scope
- Summary of Activities & Accomplishments
  - Data Stewardship
  - Process & Strategies
  - Semantic Web
  - Services, Interoperability, and Orchestration
- Breakout Session Agenda

Tech Infusion Working Group

- Mission
  - Enable NASA's Earth Science community to reach its research, application, and education goals more quickly and cost effectively through widespread adoption of key emerging information technologies
- Scope
  - Information technologies that...
    - Provide capabilities critical to the ESD mission & vision
    - Have been substantially developed (TRL6-9) but have not been widely deployed
    - Cannot be obtained simply through reuse of mature subsystems or software
    - May be slow to adopt because of the unique characteristics of Earth science (e.g., high data volumes)
- Approach
  - Improve community understanding of the technology infusion process
  - Identify barriers and solutions to technology adoption
  - Use case studies to evaluate effectiveness of infusion processes
  - Identify and evaluate new and emerging technologies
  - Develop roadmaps for adoption of key technologies

Technology Infusion Process



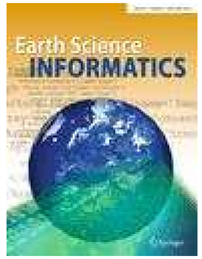
TIWG 2011 Activities

- Maintained 4 active subgroups areas
  - Data Stewardship
    - Ruth Duerr (NSIDC), Curt Tilmes (GSFC)
  - Infusion Process and Strategies
    - Steve Olding (GSFC/Everware-CBDI), Joe Glassy (Lupine Logic/UMT)
  - Semantic Web
    - Brian Wilson (JPL), Mike Little (LARC) / Hook Hua (JPL)
  - Services, Interoperability, and Orchestration
    - Hook Hua (JPL), Anne Wilson (LASP)
- Conducted weekly telecons
  - Data Stewardship and Full Working Group (1st Thursday)
  - Process and Strategies (2nd Thursday)
  - Semantic Web (3rd Thursday)
  - Services Interoperability (4th Thursday)

Data Stewardship

Data Stewardship Subgroup

- Completed identifiers paper
  - "On the utility of identification schemes for digital earth science data: an assessment and recommendations“
  - Ruth E. Duerr, Robert R. Downs, Curt Tilmes, Bruce Barkstrom, W. Christopher Lenhardt, Joseph Glassy, Luis E. Bermudez and Peter Slaughter
  - DOI: 10.1007/s12145-011-0083-6
  - http://www.springerlink.com/content/52760gq3h200gw38/



- Submitting UUID and DOIs to SPG
  - Original plan to submit as standards but need 2 in-NASA ES implementations to get on the standards track
  - UUIDs to be submitted as tech note (Joe Glassy working on, based on Measures work)
  - DOI submission possibly superseded by John Moses’ work to have NASA generate DOIs.
- Developed data lifecycle technology hype cycles
  - Identifier technologies
  - Metadata technologies
  - Data format technologies
  - Data stewardship technologies
  - Repository technologies

Hype Cycles Phases

- "Technology Trigger"
  - The first phase of a Hype Cycle is the "technology trigger" or breakthrough, product launch or other event that generates significant press and interest.
- "Peak of Inflated Expectations"
  - In the next phase, a frenzy of publicity typically generates over-enthusiasm and unrealistic expectations. There may be some successful applications of a technology, but there are typically more failures.
- "Trough of Disillusionment"
  - Technologies enter the "trough of disillusionment" because they fail to meet expectations and quickly become unfashionable. Consequently, the press usually abandons the topic and the technology.
  - Less visibility. More users.
- "Slope of Enlightenment"
  - Although the press may have stopped covering the technology, some businesses continue through the "slope of enlightenment" and experiment to understand the benefits and practical application of the technology.
- "Plateau of Productivity"
  - A technology reaches the "plateau of productivity" as the benefits of it become widely demonstrated and accepted. The technology becomes increasingly stable and evolves in second and third generations. The final height of the plateau varies according to whether the technology is broadly applicable or benefits only a niche market

Placement on the NASA Earth Science Hype Cycle

- Maturity
  - Technical completeness / operational readiness
  - 'Widespread' use in NASA production applications
- Visibility (examples)
  - References in conference presentations / papers
  - Number of press / journal articles
  - Inclusions in NASA proposals for funding
- Rate of progression
  - How to estimate projected rate of adoption?

Estimated years to mainstream adoption

- ○ < 2 years
- ○ 2-5 years
- ● 5-10 years
- ● > 10 years
- ● Obsolete before plateau
- ● deprecated

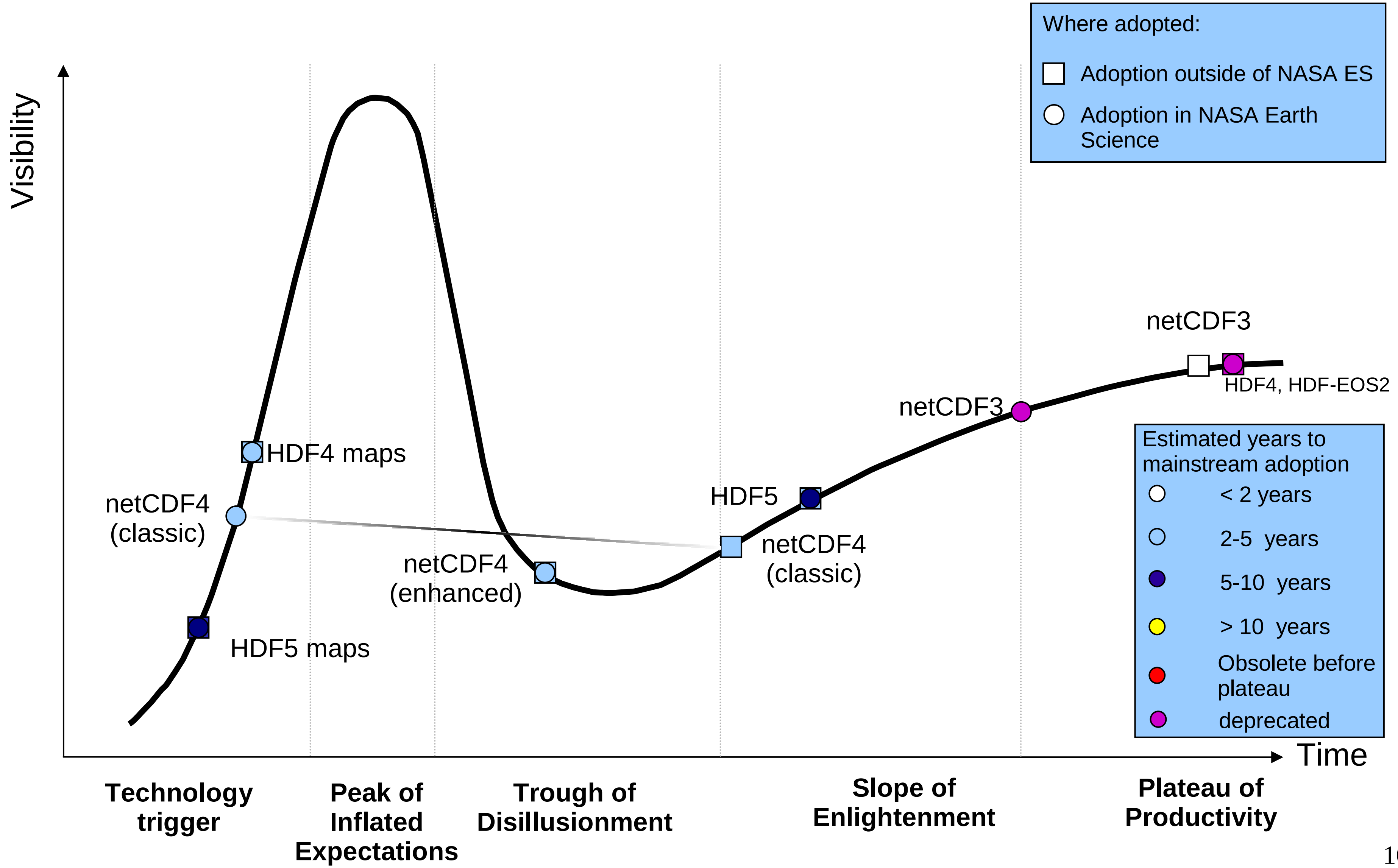
- All of these as if they ended with “in the Earth Sciences”

Where adopted:

- □ Adoption outside of NASA ES
- ○ Adoption in NASA Earth Science

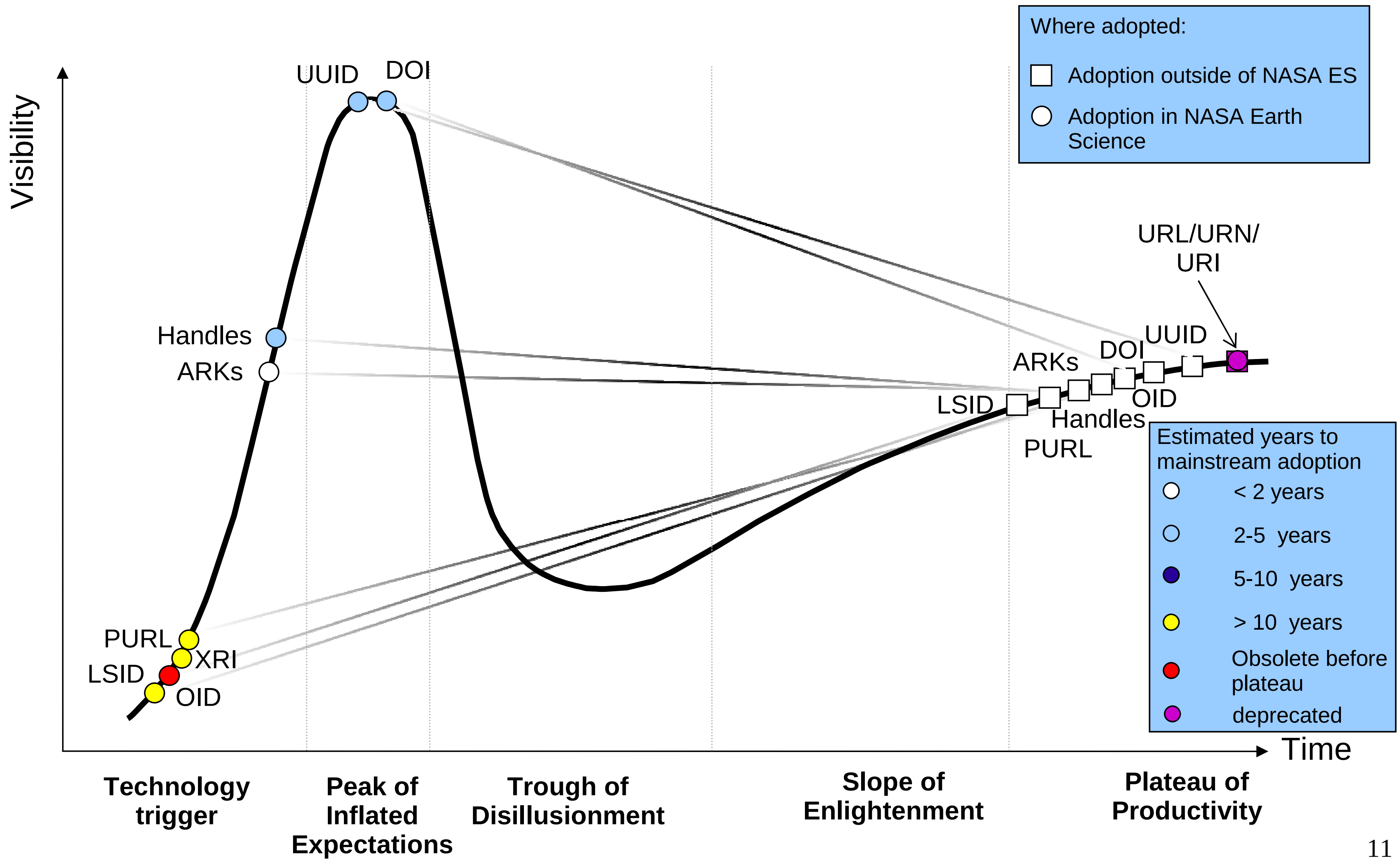
2011 Hype Cycle for Emerging Data Format Technologies

Based on the Gartner Hype Cycle methodology, adapted by the Technology Infusion Working Group for the Earth Science community



2011 Hype Cycle for Emerging Identifier Technologies

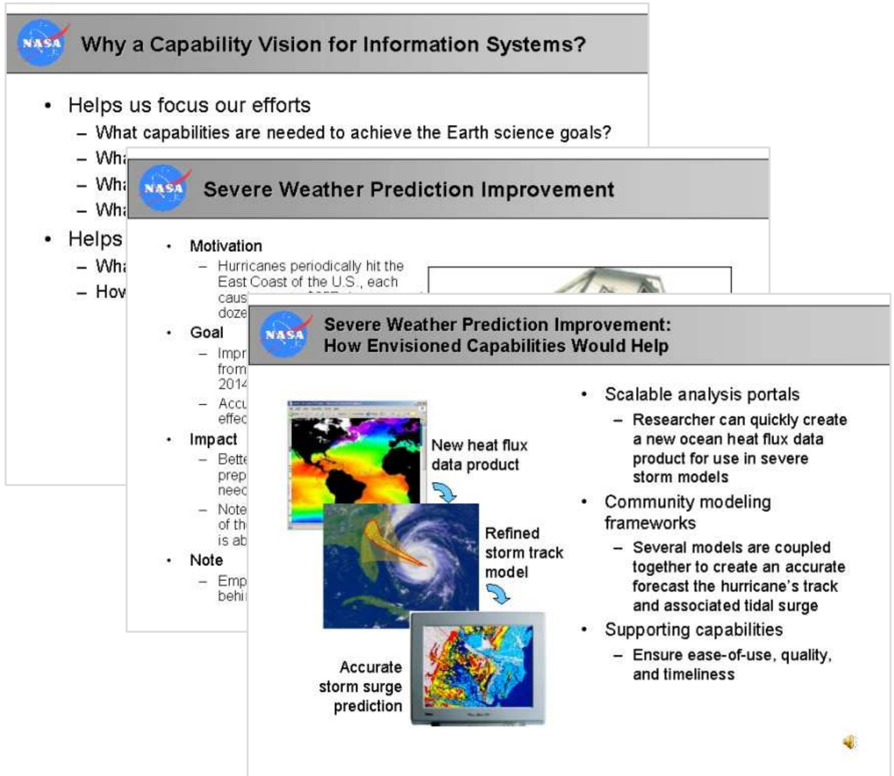
Based on the Gartner Hype Cycle methodology, adapted by the Technology Infusion Working Group for the Earth Science community

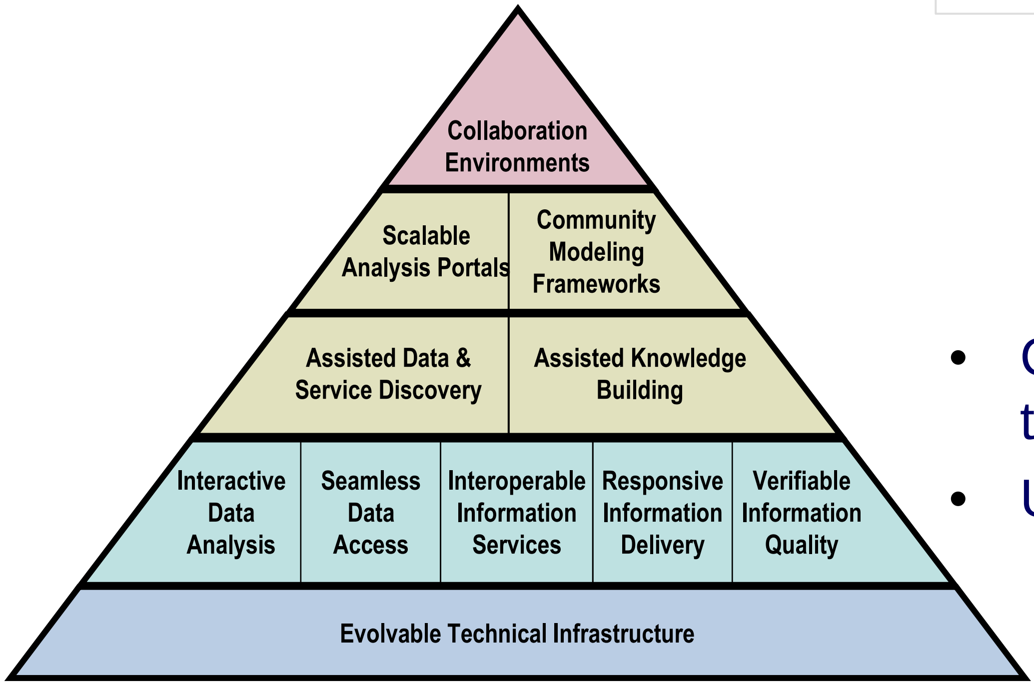


Process and Strategies

Capability Vision for Earth Science Data Systems

- Describes 11 high-level capabilities comprising an Earth science information system capability vision
- Identifies technologies critical to achieving the vision
- Positions capabilities and technologies within a real-world scenario
- Used to develop a shared understanding of the vision within the community





- Continue to review and maintain the Capability Vision
- Updated
  - Collaboration Environments
  - Integrated Analysis Portals

Recommendations for Data Level Interoperability

- This document defines a set of pragmatic recommendations to:
  - Achieve best data interoperability.
  - Promote usability of the data.
  - Promote data transparency.
  - Promote consistency and reliability in how users identify and cite earth science datasets.
- Recommendations
  - Data and file formats
  - File names
  - Identifiers
  - Metadata
- Will be made available on earthdata.nasa.gov

TIWG Recommendations for Data Level Interoperability (2010-2011)

Goals/Objectives

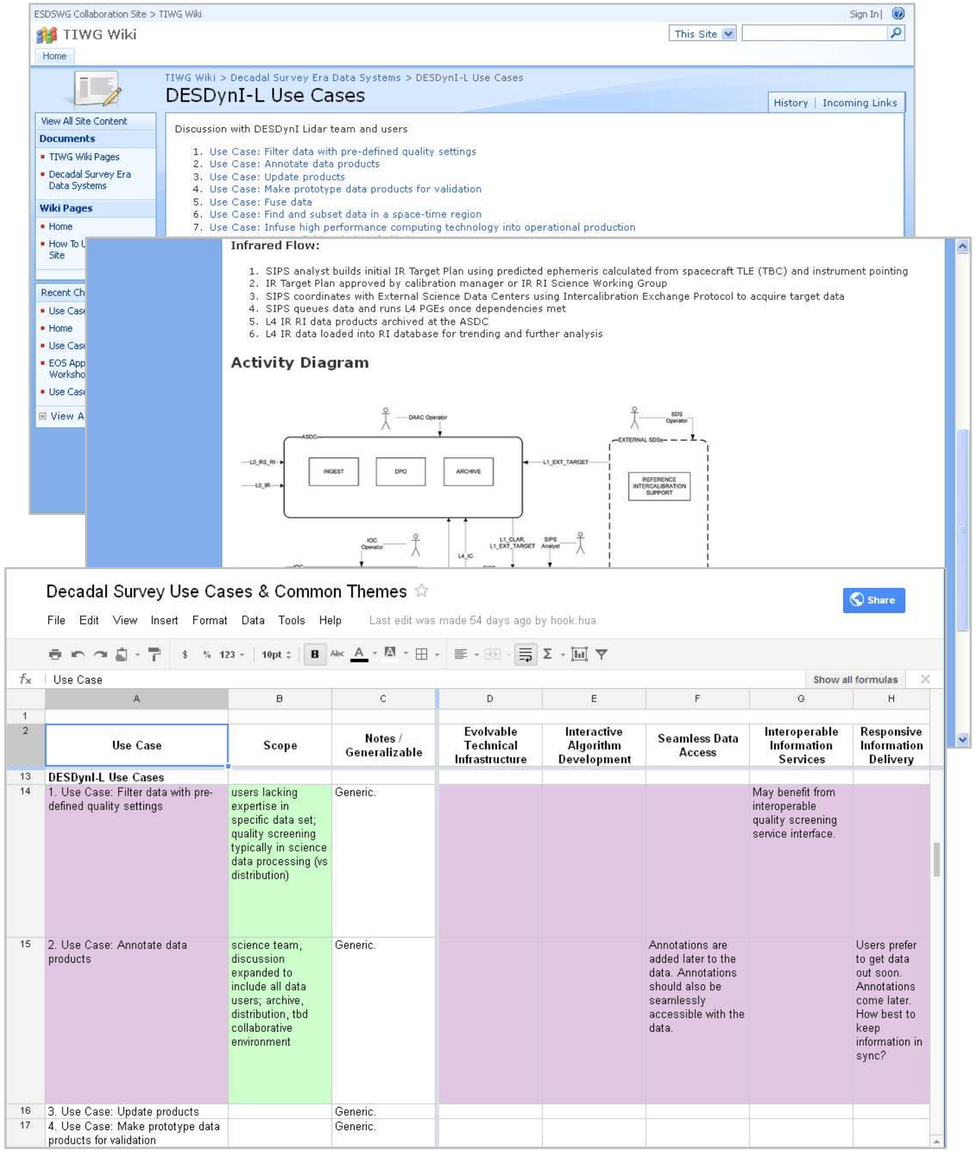
Data Level interoperability involves an on-going interaction (and conversation) between Earth data producers and consumers. Ideally the practice of interoperability should include the perspective and feedback of data consumers and producers alike, not just data producers. In a NASA Decadal Survey and NASA Measures context, team members and PI's frequently occupy a mix of both roles. For example, while primarily concerned with producing a NASA Measures dataset, a typical team frequently also uses data as upstream inputs to a production process and therefore experiences firsthand both the benefits (or shortcomings) of any interoperability practices in use.

This document defines a set of pragmatic recommendations to:

- achieve best data interoperability.
- promote usability of the data. The ESDSWG provides recommendations on technologies and conventions (frequently in support of formal standards) to enhance the usability of the data we produce, as well as the services used to identify, discover, access, and process the data.
- promote data transparency , emphasizing the use of unique identifiers, and data provenance practices.
- promote consistency and reliability in how users identify and cite earth science datasets (via unique identifiers), as well as in how datasets are discovered , accessed and retrieved. For new and experienced users alike, the experience of working with Earth science data should be predictable, consistent, and ideally -- repeatable -- across diverse types and sources of data.

Decadal Survey Mission Use Cases

- Past activities
  - Capture use cases for DS missions
  - Analyze use cases and align with the Capability Vision
  - Breakout session at 2010 summer ESIP meeting
    - TIWG DS Mission Use Case Workshop
    - Start to identify common themes and technology needs.
- Current activities
  - Review use cases for general applicability as DS missions reprioritized.
  - Summer ESIP meeting - Technology Infusion for the Decadal Survey Era: Data Quality Capability Needs
  - Technology Infusion Analysis of DS Era Use Cases
    - Identify technology gaps and infusion opportunities
    - SIO subgroup: Seamless Data Access, Interoperable Information Services, and Collaborative Frameworks / Environments



Semantic Web

Semantic Web

- Semantic Web Subgroup established in 2006.
- Focus has evolved as community understanding has matured
  - Raising awareness of semantic web
  - Education on core semantic concepts and technologies
  - Practical applications of semantic technologies
- Delivered tutorials and workshops at ESDSWG and ESIP Federation meetings
  - Introduction to Semantic Web
  - Semantic Query, Rules and Knowledge Encoding
  - Semantic Web Technology
  - Practical aspects of creating semantic web applications
  - Hands-on tutorial for developing a semantic web application (Presented at the January 2011 ESIP meeting)
  - *Provenance Ontologies (ESIP 2011 Winter Meeting)*
  - *Data Life-Cycle Provenance Ontology (ESIP 2011 Summer Meeting, Joint session with Data Stewardship and Preservation Cluster)*
- Developed technology roadmaps and hype cycles to assess and track maturity of semantic technologies.
- Developing demonstrations of semantic technologies from initial proof of concept through to practical applications
  - Planned for January 2012 ESIP meeting.

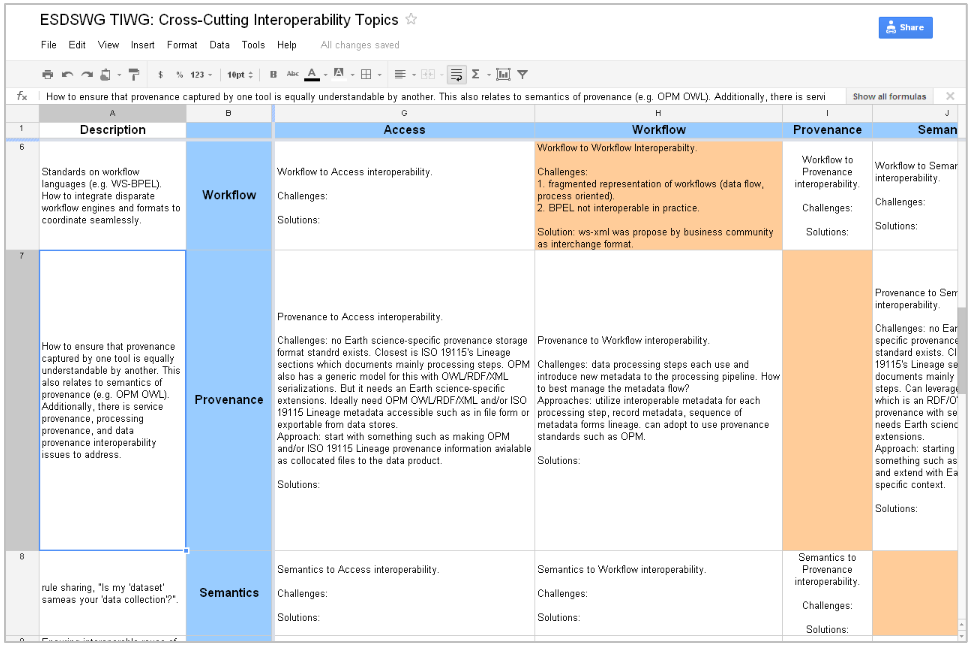
Current Projects Using Semantic Web Technologies

- Data Quality Screening Service (DQSS) – ACCESS, C. Lynnes
  - Using ontology to select quality flags by dataset and apply expert-driven quality thresholds to satellite datasets, pixel by pixel
- Provenance Services for a MEASUREs product – Multi-Sensor Water Vapor Climatology Stratified by Cloud Classes, ACCESS, H. Hua
  - Capturing production provenance as logical triples in RDF format, using the Open Provenance Model (OPM) OWL ontology for interoperability
  - Provenance handling using SPARQL, Rules, and faceted search.
- Noesis 2.0 Smart Meta-Search – R. Ramachandran, part of Service & Event Casting ACCESS project (B. Wilson)
  - Search for relevant datasets & services across multiple back-ends: GCMD, ECHO, casts, Google, news, etc.
  - Uses ontology to expand query terms: broaden, narrow, synonyms
- Linked Open Research Data for Earth Science Informatics – Funding Friday project, E. Rozell, T. Narock
  - Mining AGU abstracts for links between people, papers, topics, sessions
  - Publish as Linked Open Data (LOD) on web in RDF/OWL format
- ESIP Federation People, Skills, Collaboration Database – E. Robinson
  - Simple ontology for people/skills/collab.; will be linked to AGU data

Services, Interoperability, and Orchestration

Cross-cutting Interoperability

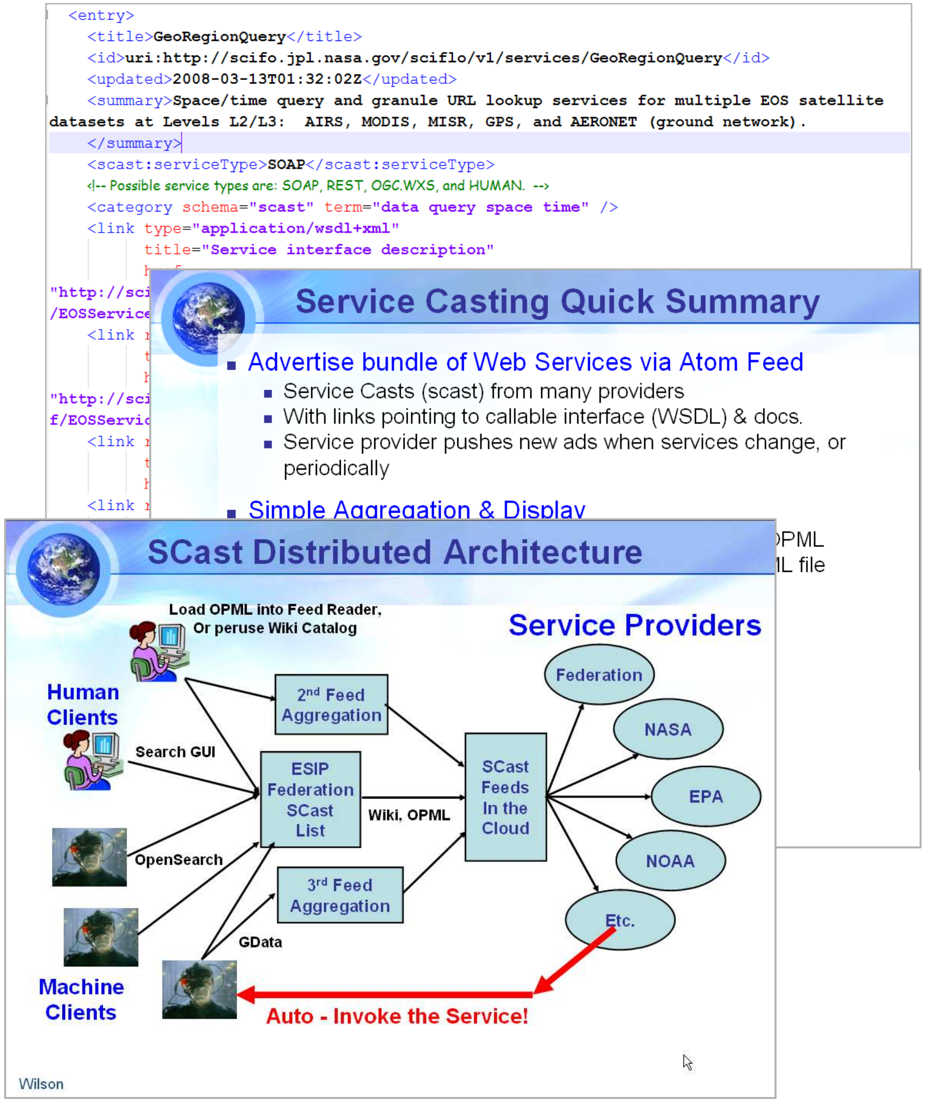
- Analysis of challenges and possible solutions for complex interoperability issues
- Various interoperability dimensions:
  - Data, Metadata, Discovery, Access, Workflow, Provenance, Semantics, Infrastructure



- Example:
  - Provenance to Discovery Interoperability
- Challenges:
  - Discovery based on interoperability of metadata.
  - ISO has no place for file identifier, was suggested to put into free text.
  - Too many provenance encodings: ISO, OPM, PML, W3C Provenance
- Solutions:
  - W3C Provenance is developing a provenance interchange format. In a domain-neutral way. Takes a more open-world approach so that domain models can be made to fit. ISO has more constrained approach for domain adaptation. Has danger of becoming the overarching format.

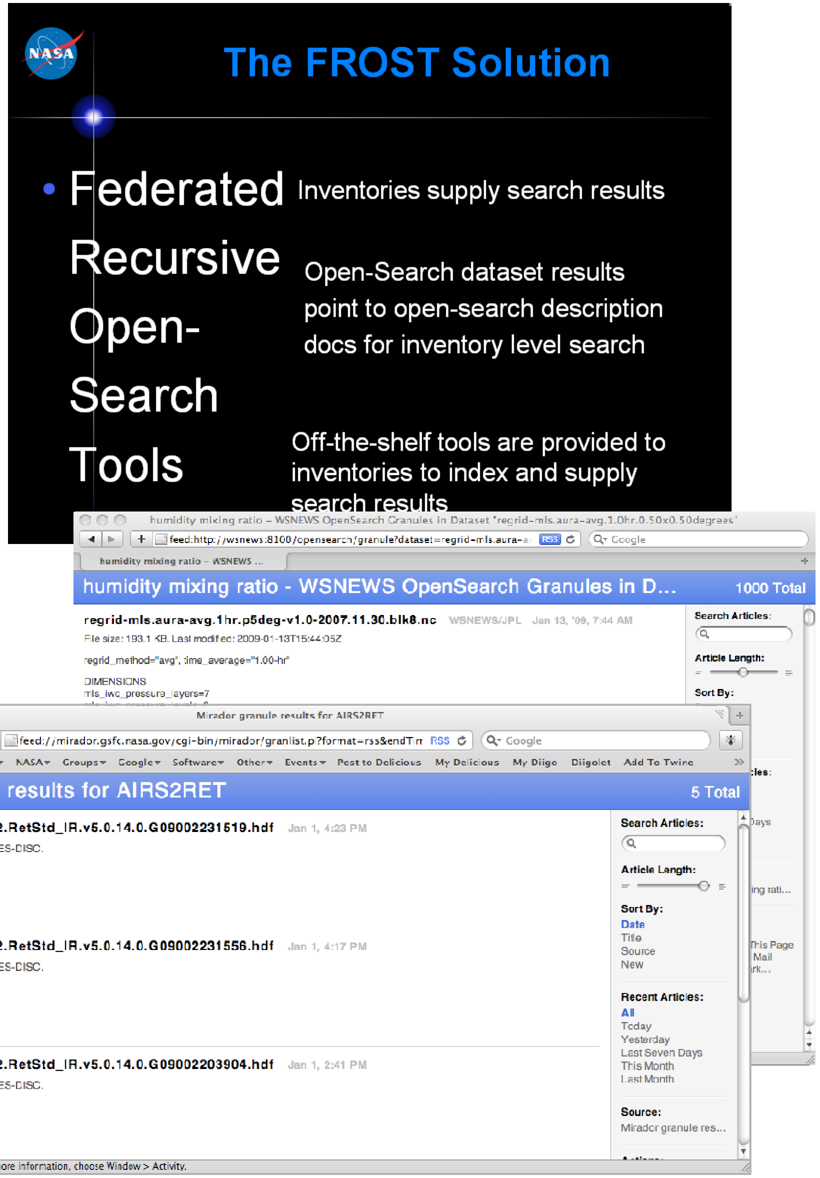
Data and Services Discovery: Casting

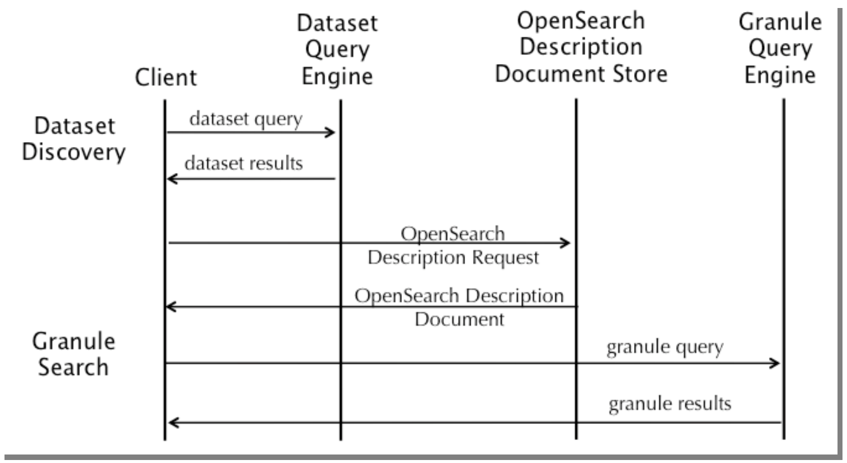
- Data and Service Casting
  - Lightweight, distributed alternative to centralized directories/registries
  - Easy authoring of advertisements using Atom syndication feeds
  - Providers re-publish feeds when data or services change (provider in control)
  - Scalable discovery, no single registry
  - Original service casting concept created by Brian Wilson in response to a TIWG assessment of web services registries
  - “Extended” to include granule level data casting (JPL) and collection level data casting (Ruth Duerr)



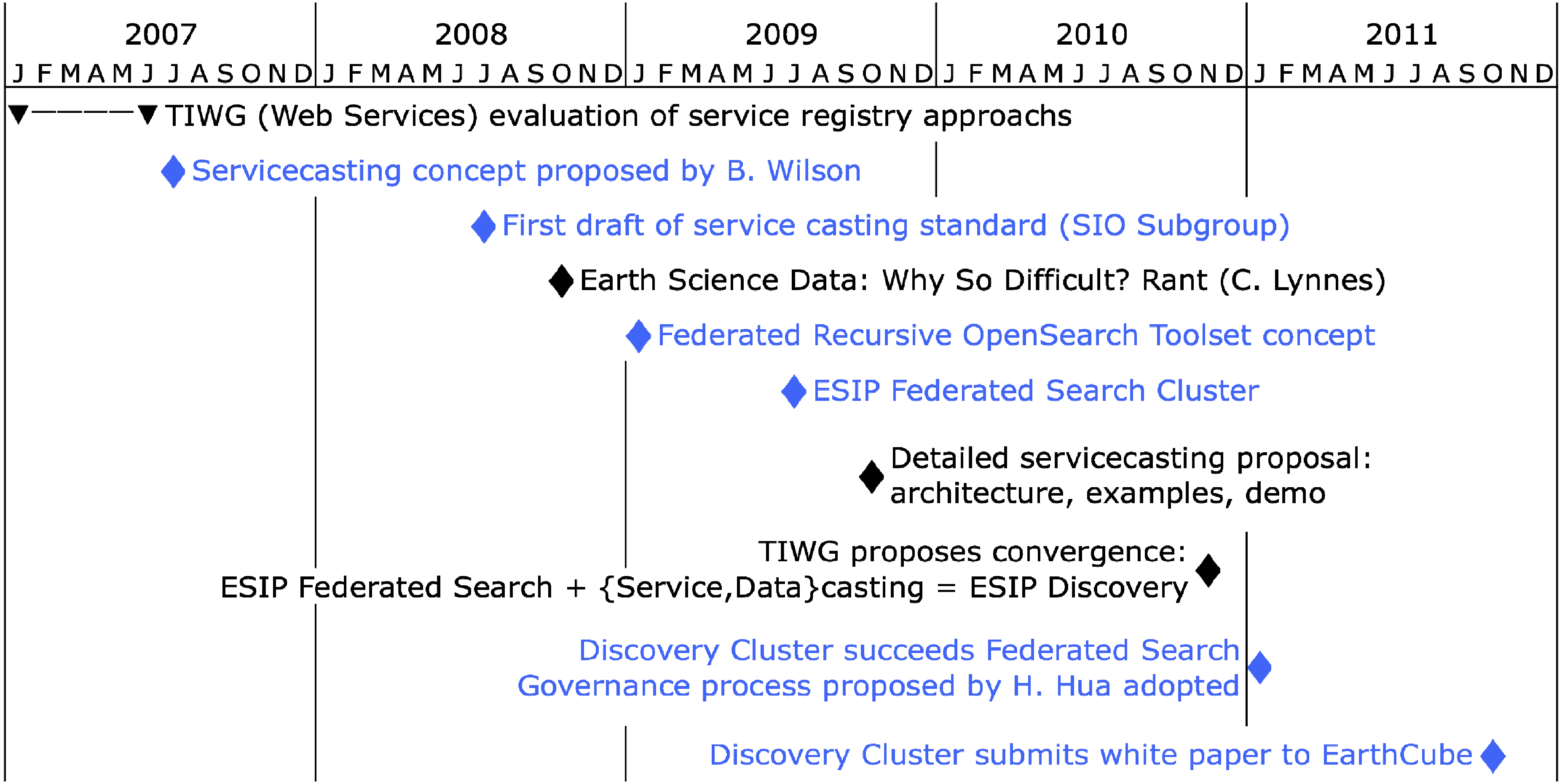
Data and Services Discovery: OpenSearch

- Federated OpenSearch
  - OpenSearch: search request/response conventions
    - Time & Space extensions
  - Simple, lightweight, widespread, extensible.





Discovery History



Data and Services Discovery: Today

- Multiple Discovery implementations across different centers
  - ESIP Federated OpenSearch Services
    - GSFC, JPL, ECHO, NSIDC, UAH, GHRC, MODAPS/LAADS, WS-NEWS
  - DataCasting Services
    - JPL PO DAAC, NSIDC, UAH
  - ServiceCasting Services
    - JPL and NSIDC
  - Discovery Clients
    - GES DISC, JPL, EOSDIS, NSIDC, UAH, OMI SIPS.
- Governance process being exercised
  - DCP-1: *ESIP Discovery Cast Atom Response Format v1.1*
  - DCP-2: *Canonicalizing Granule-level OPeNDAP Links (withdrawn)*
  - DCP-3: *Standardized Linking from one cast to another or to additional metadata*
- Funded Discovery Testbed activity by ESIP Products &Services Committee
- ESIP Discovery playing key role in *Earth Science Collaboratory* vision
  - Contributed white paper to NSF EarthCube

- AGU
  - Fall 2010 AGU Meeting
    - Interoperability Barriers for Earth Science Data Systems (Hook Hua and Anne Wilson session co-conveners)
    - Information Technology Infusion Success Strategies (Brian Wilson session co-convener)
    - A Toolbox for Organization-wide Infusion of Data Systems Technologies (TIWG poster)
  - Fall 2011 Meeting
    - Interoperability Solutions in Earth Science Data Systems (Hook Hua session co-convener)

TIWG Breakout Agenda

- #1 Planning (Tue 3:40 – 5:30 pm)
  - Follow up to working groups restructure plan and ESDIS technology directions
- #2 Semantic Web (Wed 8:30 – 10:35 am)
  - Semantic web intro (20 mins)
  - Semantic web technology stack. Future tutorials and the testbed activities. (30 mins)
  - 2012 planning; organizing demos, modeling and ontology development, strategies for moving forward with semantic web. (40 mins)
  - Discovery intro; data and service casting standards (30 mins)
- #3 Data Stewardship (Wed 10:50-12:15 pm)
  - NASA Earth Science Data Preservation Content Specification. Rama. (20 mins)
  - What is NASA doing about identifiers and citations? (20 mins)
  - 2012 Planning session (30 mins)

TIWG Breakout Agenda

- #4 Interoperability (Wed 3:50pm - 5:30pm)
  - SIO review (20 mins)
  - Interoperability pain points - rants (30 mins).
  - 2012 planning (30 mins)
  - Splinter groups (20 mins)
    - Discussion and planning for discovery
    - Taxonomy of technologies, standards, and best practices
    - Future planning
- #5 Wrap up (Thur 8:30am – 9:30am)
  - Tech talk: webification / pomegranate (15 mins)
  - Breakout session and SIO splinter group conclusions
  - Final planning